Aspects of Risk Adjustment in Healthcare

1. Exploring the relationship between demographic and clinical risk adjustment
   Matthew Myers, Barry Childs

2. Application of various statistical measures to hospital case mix adjustment
   Shirley Ginsberg, Barry Childs
Exploring the relationship between demographic and clinical risk adjustment

Matthew Myers - The Health Monitor Company
Barry Childs - Lighthouse Actuarial Consulting
Agenda

- Risk Adjustment
- Risk Adjustment in healthcare
- Healthcare data
- Clinical risk adjustment
- Demand and supply data for prediction
- Demographic vs. Clinical factors
- A Case study – measuring primary care practitioner efficiency
Risk Adjustment

• Identification of factors whereby individuals with same profile have a homogeneous claim profile, e.g.
  – Car insurance – over / under age 25
  – Life insurance – smoker / non-smoker
  – Health insurance – chronic / non-chronic medicine user

• Uses of risk adjustment
  – Risk rating – Charge appropriate premium for specific risk
  – Community rating – understand impact of change in risk profile
Risk Adjustment in healthcare

• Key features of SA healthcare environment
  – Open enrolment
  – Community rating
  – No compulsory membership

• Risk adjustment techniques are imperative in understanding underlying risk profile

• Understanding of how a change in risk profile impacts on demand / cost of provision of care
Healthcare data

- Healthcare provides credible data – high claims frequency compared to other insurance lines
- SA private healthcare industry is primarily Fee for Service based → rich body of claims data available
  - Demographic data (age, gender, location, chronic status, benefit option)
  - Claims data (practice codes, procedure code, diagnosis code (ICD-10), tariff and benefit amounts)
- Introduction of ARMs may result in less accurate data being collected.
Clinical Risk Adjustment

- Various techniques available
  - DRGs
  - Patient risk adjustment
  - Clinical episode groupers

- Challenges
  - Availability of data
  - At short duration, no data exists!
  - Quality of coding
Risk rating data

- Diagnosis / procedure codes
- Option choice
- Chronic Status
- Age, gender, location

Clinical data – harder to collect and analyse in a meaningful way

Demographic data – readily available
## Factors to consider in risk adjustment

<table>
<thead>
<tr>
<th>Considerations</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic policy holder information</strong></td>
<td>Age, gender, chronic status, cover type, region, industry, individual / corporate membership</td>
</tr>
<tr>
<td>Population demographics determine basic level of demand</td>
<td></td>
</tr>
<tr>
<td><strong>Diagnostic information</strong></td>
<td>Hypertension, Diabetes, HIV, Asthma…</td>
</tr>
<tr>
<td>Disease profiles give more granular insight into aspects of demand</td>
<td></td>
</tr>
<tr>
<td><strong>Healthcare services</strong></td>
<td>Hospital admissions, pathology tests, doctor visits</td>
</tr>
<tr>
<td>Actual services sought match cost but can impair certain purposes such as profiling</td>
<td></td>
</tr>
</tbody>
</table>
Using Clinical data

• Data is not always as clean and accurate as demographic data
• Key is to be able to meaningfully summarise the amount of clinical information
• We have used Johns Hopkins’ Adjusted Clinical Groups (ACGs)
  – System uses clinical data to group patients into homogeneous clinical and resource intense groups
  – Provides means of evaluating population health, allowing risk profile analysis retro- and pro-spectively
ACG Uses

- Performance Assessment
  - Variations in profile mix, productivity and utilization
  - Payment differentiation
  - Fraud and abuse investigation
- Health Status Monitoring
- Care Management
  - High Risk Case Identification
  - Disease Management Case-mix control
- Finance
  - Payment/Rate Setting/Underwriting
- Provider profiling – explored later in this talk
Model fitting and value judgment

• The conventional measure for model performance is the $R^2$ statistic
• We are seeking to remove as much of the variance we can explain away, while leaving the underlying natural variation in place
• So, a poor $R^2$ could indicate a poor model performance, or just high underlying variation
• When using models to profile providers of care, this becomes a value judgment
## \( R^2 \) Results using different risk factors

<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age Only</td>
<td>3.4%</td>
</tr>
<tr>
<td>Age and Gender</td>
<td>3.6%</td>
</tr>
<tr>
<td>Age, Gender and Chronic Status</td>
<td>4.3%</td>
</tr>
<tr>
<td>Age, Gender, Chronic, &amp; Option</td>
<td>5.5%</td>
</tr>
<tr>
<td>ACG only</td>
<td>15.0%</td>
</tr>
<tr>
<td>ACG and Option</td>
<td>15.8%</td>
</tr>
</tbody>
</table>

*Analysis based on large restricted medical scheme, 157,000 beneficiaries*
Choice of a clinical grouper

- The Society of Actuaries periodically conducts a review of patient classification systems
- Some results from their latest review:

<table>
<thead>
<tr>
<th>Risk Adjuster Tool</th>
<th>Developer</th>
<th>R Sq</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACGs</td>
<td>Johns Hopkins</td>
<td>19.20%</td>
</tr>
<tr>
<td>CDPS</td>
<td>Kronick / UCSD</td>
<td>14.90%</td>
</tr>
<tr>
<td>Clinical Risk Groups</td>
<td>3M</td>
<td>17.50%</td>
</tr>
<tr>
<td>DxCG DCG</td>
<td>DxCG</td>
<td>20.60%</td>
</tr>
</tbody>
</table>
Relationship between the two approaches

- Correlation Coefficient between method predictions is 40.8%
- Correlation Coefficient of the residual error term for each method is 93.7% - resulting from low fit overall for both models
- Which *should* mean that the residual variation is not explained by risk profile, but some other underlying cause, such as variation in provider practice
Alternatives to $R^2$

- **AIC** - Akaike's information criterion
- Penalises the introduction of additional parameters
- Ranks measures to compare model performance (lower is better)

<table>
<thead>
<tr>
<th>Alternatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age only</td>
</tr>
<tr>
<td>Age, Gender</td>
</tr>
<tr>
<td>Age, Gender, Chronic status, Option</td>
</tr>
<tr>
<td>ACG</td>
</tr>
<tr>
<td>ACG and Option</td>
</tr>
</tbody>
</table>
A Case Study: Measuring primary care provider efficiency

Risk Adjustment is a critical part of the profiling process.
Measuring primary care provider efficiency using demographics

Demographic Factors

Actual - Expected Costs plpm

Months Exposed for allocated patients
Measuring primary care provider efficiency using clinical data

ACG Factors

95% Confidence Interval

Actual - Expected Costs plpm

Months Exposed for allocated patients
The relationship between the methods

A-E Demographic Risk adjustment vs. A-E using ACG Risk adjustment

‘double’ confirmation that these doctors cost more than their peers
Other benefits of using clinical data and grouper

• Easier to include quality metrics
• Gain more insight into moving parts – more specific information on what is driving costs
• Easier to share meaningful information with the providers of care
  – Share relevant clinical information with providers
  – Not that average age increased by one, but that there were more Congestive Heart Failures, and/or that doctors A’s cost of treating Hypertension compared to Doctor B’s.
In conclusion

• **Clinical data** provides a much better basis for risk adjustment than demographics
  – The fit is better
  – The interpretation of the results is far better – gain insight into **disease profile** and cost weights of each disease in the population
• Clinical methods present some challenges for projection purposes – for example, the profile of clinical claims is not available in advance
• Clinical approach far better for profiling
Application of various statistical measures to hospital case mix adjustment

Shirley Ginsberg - Discovery Health
Barry Childs - Lighthouse Actuarial Consulting
Agenda

- Contextualising hospital costs in South Africa
- Statistical models for hospital cost analysis
- Diagnosis Related Groups
- Is there a case for using further risk adjustment factors?
- Trimming methodologies analysed
- Conclusions
Contextualising in-hospital costs in South Africa

Council for Medical Schemes Annual Reports 1980 - 2008

Primary Care
Hospital
Specialists
Medicines

Year

Cost per person in 2008 prices
0 500 1,000 1,500 2,000 2,500 3,000 3,500
**Statistical models for hospital cost analysis**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Government funder</strong></td>
<td>Enables effective budgeting, planning, resource allocation, infrastructure planning</td>
</tr>
<tr>
<td></td>
<td>Cost, Quality and Value trend analysis to inform health policy</td>
</tr>
<tr>
<td><strong>Health Insurer</strong></td>
<td>Cost, Quality and Value trend analysis to inform benefit design and claims experience analysis</td>
</tr>
<tr>
<td></td>
<td>Informs price negotiations, hospital network designs and pricing, and managed care policy</td>
</tr>
<tr>
<td><strong>Hospitals</strong></td>
<td>Trend analysis and identification of areas and sources of inefficiency</td>
</tr>
<tr>
<td></td>
<td>Profitability analysis with respect to region, case types, specialities</td>
</tr>
</tbody>
</table>
Diagnosis Related Groups (DRGs)

“the single most influential postwar innovation in medical financing”

- DRGs are a hybrid Clinical/Statistical derived algorithm intended to group admissions into a manageable and statistically significant number of clinically similar groups that are expected to have similar resource use.
- Increasing use around the world for hospital payment.
- Key application is measurement of Case Mix.
- Conceptually closest to a Cluster Model, with clinical differentiators.
- Key issue is demand factors versus supply factors.

Wealth of literature and research – contact authors for history of DRGs, applications of models, technical issues, etc.
Structure of Discovery Health’s DRG grouper

Clinical and Demographic Data

Major Diagnosis Groups – body system level

Clinically homogenous groups with similar resource use

Base DRGs – Medical / Surgical / Obstetric Classification, as well as diagnostic and procedure description

Statistical models for complication and co-morbidity differentiation

DRGs – base DRGs differentiated into higher resource bands based on clinical codes
Example of a case mix calculation

The impact of adjusting by case mix

Average admission cost

Case Mix Index

- Average Cost
- Average cost case mix adjusted
- Case Mix index

Hospital A

Hospital B
Can we improve on the DRG?

• Are there other factors of a patient that influence the patient’s *demand* for healthcare that DRGs do not take into account?

• There are some tradeoffs
  - *More cells means scantier data*
  - *Difference between demand and supply factors and the impact of incorporating supply side factors in the model*
  - *Correlation vs. causation argument of parameter estimates*
  - *Parameter estimates cannot be interpreted independently of the factors fitted in the model*
Case Study

- Can DRGs enable us to price a group of lives onto a new benefit pool without making assumptions of plan choice?
Consideration of GLM as an additional layer

• The following demographic factors are used:
  - Age band
  - Chronic condition count
  - Province
  - Gender
  - Plan code

• Three GLMs were analysed:
  - demographic factors
  - demographic factors and DRGs
  - demographic factors excluding plan and DRGs
Results

• Adding DRGs improves on the demographic model

• Dropping plan has minimal impact
Results

Mean absolute prediction error of hospital cost
Mean absolute under prediction error of hospital cost
Mean absolute over prediction error of hospital cost
R-squared
% of cases under predicted
% of cases over predicted
The Case for Trimming

There are some shortcomings in the methods:

- Categories not strictly homogeneous
- Clinical coding is not perfect and cause misallocation
- Expected values based on mean costs, and therefore sensitive to outliers
- Costs are often very right skewed
- Hospitals are not insurers (not here anyway)

Identification of outliers and trimming is required to improve our understanding of the tail and the impact of abnormal cases
The Case for Trimming

- Each DRG contains admissions which have a particular cost distribution
- The outlier identification process is intended to identify admissions that are non-typical

Low outliers due to:
- early patient discharge
- incorrect coding
- data errors

Upper outliers due to:
- incorrect coding
- data errors
- unusual costs incurred
Some other trimming issues to consider

• Aim to identify outliers in a *consistent* manner
• Avoid the onerous and time consuming task of audits
• Audits are only possible when you have the detailed data which some entities will not in position to evaluate
• DRG groups with low admission counts a different problem
• Consider *truncation* as an alternative
• *Tradeoff* between $R^2$ and loss of data
Behaviour of the underlying data

- Very detailed hospital cost data available in South Africa
- Most of the literature uses length of stay as surrogate measure for resource use, and trimming methods built on this practice.
- 91% of DRGs containing 99.8% of admission data exhibit right skewed distributions
- It is the right tail that contributes mostly to the variability
- Therefore the more events that are trimmed from the right tail of each DRG’s cost distribution, the greater the improvement in R-squared.
Trimming methodologies analysed

Methods considered:
• The inter-quartile range
• The adjusted inter-quartile range
• The adjusted inter-quartile range with truncation
• The geometric mean
• The adjusted box plot using the medcouple statistic
• The adjusted inter-quartile range using the medcouple

Measures:
• $R^2$, CV, % of Admissions and % of money trimmed out.
Illustration of trim points

<table>
<thead>
<tr>
<th>Caesar</th>
<th>Head Trauma W/O CC</th>
<th>Cesarean Delivery W MCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentile 25%</td>
<td>Percentile 75%</td>
<td>Median</td>
</tr>
<tr>
<td>Caesar</td>
<td>15109</td>
<td>18103</td>
</tr>
<tr>
<td>Head Trauma</td>
<td>2078</td>
<td>6503</td>
</tr>
</tbody>
</table>
The Inter-quartile range

Adjusted IQR
Outlier_top > Quartile_{75%} + k*(Quartile_{75%} - Quartile_{25%})
Outlier bottom < Quartile_{25%}/3.5

Symmetrical IQR
Outlier_top > Quartile_{75%} + k*(Quartile_{75%} - Quartile_{25%})
Outlier bottom < Quartile_{25%} - k*(Quartile_{75%} - Quartile_{25%})
The Geometric mean

Trimming with the geometric mean at the upper end

Provides a higher R-squared than the adjusted IQR for the same % of events trimmed.
Using the med-couple statistic for an adjusted box-plot for skewed distributions

- \( MC = \frac{\text{med}_{x_i \leq x \leq x_j} h(x_i, x_j)}{\text{median}_n} \)
- Where \( h(x_i, x_j) = \frac{(x_j - m_n) - (m_n - x_i)}{(x_j - x_i)} \) and \( m_n \) is the median of \( X_n \)
- The MC statistic lies between -1 and 1
- It is positive for right skewed distributions and negative for left skewed distributions
- This statistic is used (Mia Hubert, 2006) to create an adjusted box plot for skewed distributions
- When MC > 0, all observations outside the interval
  \[ Q1 - 1.5e^{-4MC} \text{IQR}; Q3 + 1.5e^{3MC} \text{IQR} \]
  will be marked as potential outlier.
- For MC < 0, the interval becomes
  \[ Q1 - 1.5e^{-3MC} \text{IQR}; Q3 + 1.5e^{4MC} \text{IQR} \]
Using the med-couple statistic for an adjusted box-plot for skewed distributions

- R-squared values much lower
- Method is not intuitive - it trims out less from more right skewed distributions that from distributions less skewed from the right
- Results in including more variability at each DRG level and hence having a lower resulting R-squared value

<table>
<thead>
<tr>
<th>K</th>
<th>1</th>
<th>1.1</th>
<th>1.2</th>
<th>1.3</th>
<th>1.4</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-Squared</td>
<td>68.8%</td>
<td>67.8%</td>
<td>66.5%</td>
<td>65.6%</td>
<td>64.8%</td>
<td>64.0%</td>
</tr>
<tr>
<td>% of money remaining after trimming</td>
<td>84.8%</td>
<td>86.2%</td>
<td>87.5%</td>
<td>88.5%</td>
<td>89.5%</td>
<td>90.4%</td>
</tr>
<tr>
<td>% of events trimmed</td>
<td>10.5%</td>
<td>9.1%</td>
<td>7.9%</td>
<td>6.9%</td>
<td>6.1%</td>
<td>5.3%</td>
</tr>
<tr>
<td>% of events defined as top outliers</td>
<td>3.9%</td>
<td>3.4%</td>
<td>3.1%</td>
<td>2.7%</td>
<td>2.4%</td>
<td>2.2%</td>
</tr>
<tr>
<td>% of events defined as bottom outliers</td>
<td>6.6%</td>
<td>5.7%</td>
<td>4.9%</td>
<td>4.2%</td>
<td>3.6%</td>
<td>3.1%</td>
</tr>
</tbody>
</table>
Adjusted IQR using the Medcouple

- Outliers at the bottom are defined to be less than
- Rationale for using this statistic- want to trim less from the lower end of right skewed distributions compared to left skewed distributions

<table>
<thead>
<tr>
<th>K</th>
<th>2</th>
<th>2.1</th>
<th>2.2</th>
<th>2.3</th>
<th>2.4</th>
<th>2.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-Squared</td>
<td>72.7%</td>
<td>72.3%</td>
<td>72.0%</td>
<td>71.6%</td>
<td>71.1%</td>
<td>70.8%</td>
</tr>
<tr>
<td>% of money remaining after trimming</td>
<td>83.8%</td>
<td>84.4%</td>
<td>85.0%</td>
<td>85.5%</td>
<td>86.0%</td>
<td>86.4%</td>
</tr>
<tr>
<td>% of events trimmed</td>
<td>4.9%</td>
<td>4.6%</td>
<td>4.4%</td>
<td>4.2%</td>
<td>4.0%</td>
<td>3.8%</td>
</tr>
<tr>
<td>% of events defined as top outliers</td>
<td>4.3%</td>
<td>4.0%</td>
<td>3.8%</td>
<td>3.5%</td>
<td>3.3%</td>
<td>3.2%</td>
</tr>
<tr>
<td>% of events defined as bottom outliers</td>
<td>0.6%</td>
<td>0.6%</td>
<td>0.6%</td>
<td>0.6%</td>
<td>0.6%</td>
<td>0.6%</td>
</tr>
</tbody>
</table>

- Overall incorporating the med-couple method is an interesting case of computational overkill for marginal gain
Some other methods explored

- Applied adjusted IQR with different k’s based on skewness of DRG
  - *No significant improvement*
- Applied adjusted IQR with different k’s for different categories of DRGs
  - *Again no significant overall improvement*
Considering the methods together

Truncation and Adjusted Box plot perform worst.

*Geomean* seems to outperform, but gains are not significant

Clear wrong models, but a few right ones to choose from.
Trimming Summary

- Of all the good methods analysed, there are generally marginal difference in R-squared values for a given amount trimmed.
- This arises mainly due to the trimming out of relatively few admissions.
- Overall, it seems as though the adjusted IQR method provides the best results in terms of a trade off between intuitiveness, computational requirements, and goodness of fit.
In Conclusion

• The superiority of clinical risk adjustment over demographic risk adjustment seems clear
• What remains to be done is to continue to explore the relationship between demographics and clinical factors
• And find new applications for the tools available to solve the many complex problems facing healthcare systems
• As Actuaries we must step up and play a pivotal role in these discussions